Человек, чей рост превышает некую среднюю норму, часто наделяется всевозможными кличками (*каланча, дылда, верзила, дядя Степа* и т.п.), а сравнения, отражающие данное значение, в основе имеют образы длинных и высоких предметов, устремленных вверх: *высокий как (пожарная) каланча*, *высокий, длинный как (коломенская) верста; высокий, длинный* (а также *худой*) *как жердь*. Среди данных фразеологизмов есть несколько, представляющих интерес в страноведческом плане: *высокий, длинный как (коломенская) верста* – образован на базе фразеологизма *коломенская верста*, т.е. высокий столб, которым в XVII в. отмечались версты между Москвой и селом Коломенское; оборот *высокий как (пожарная) каланча* связан с наличием в центре каждого русского города специального высокого здания – пожарной каланчи, на которой постоянно дежурил человек, чтобы вовремя заметить начинающийся пожар.

Таким образом, компаративные фразеологические обороты отражают наиболее значимые черты внешности человека, причем анализ образов, лежащих в основе рассматриваемых фразеологизмов, показывает, что большая их часть соотносится с различными объектами живой и неживой природы. Чаще всего в качестве исходных элементов сравнения выступают названия животных, растений, веществ и материалов, природных явлений, а также некоторых мифологических существ, что может служить доказательством древнего происхождения данных оборотов. Большая часть данных фразеологических оборотов имеет прозрачную внутреннюю форму и построена на переосмыслении значения путем метафорического переноса по цвету, форме, месту расположения, качеству и т.п.

Лексикографические источники

1. **Жуков В.П. и др.** Словарь фразеологических синонимов русского языка. — М.: Русский язык, 1987.
2. **Огольцев В.М.** Устойчивые сравнения русского языка. Иллюстрированный словарь для говорящих на английском языке. — М.: Русский язык, 1984.
3. **Ожегов С.И., Шведова Н.Ю.** Толковый словарь русского языка. — М.: Азъ, 1993.
4. **Фелицына В.П., Мокиенко В.М.** Русские фразеологизмы. Лингвострановедческий словарь. Под ред. Е.М.Верещагина и В.Г.Костомарова. — М.: Русский язык, 1990.
5. **Фразеологический словарь русского языка.** Под ред. А.И. Молоткова. — М.: Русский язык, 1978.
6. **Яранцев Р.И.** Словарь-справочник по русской фразеологии. — М.: Русский язык, 1981.

METHODS OF INFORMATION COLLECTION

Коломейцева Е. М., Макеева М. Н.
Тамбовский государственный технический университет

Second language assessment entails the collection of a great variety of information about instruction (objectives, plans, and practices), students (e.g., their needs, goals, personal background, language experiences, achievement, and attitudes and feelings), teachers (e.g., their language experiences, language skills, and attitudes), and about schools (such as the school's physical and personnel resources). Different techniques are available for gathering these kinds of information. Tests are useful for collecting information about student achievement but cannot be used for collecting any of the other types of information, however. Other methods of data collection are appropriate for these. For example, classroom observation and student conferences can garner information about the strategies students might be using to read or write in the second language; dialogue journals can shed light on students' attitudes toward their learning experiences in class. School records, curriculum documents, or other instructional materials reveal facts about the physical and personnel resources of the school system as well as about the course of instruction itself.

Some procedures for collecting information for evaluation purposes are straightforward and require no special preparation, for example, examining school records or instructional materials. In these cases, one simply locates the relevant sources of information and becomes familiar with their contents. Other methods, such as portfolios, conferences, or questionnaires, require advanced preparation and somewhat specialized procedures.

One determines which method or combination of methods is most appropriate for making a particular decision at a given time. Moreover, not all methods are suitable for collecting all types of information that you might need for evaluation purposes, as we noted earlier. Some of the methods under consideration (for example, conferences and journals) can be used for instructional as well as evaluation purposes. Teachers will, therefore, want to use a variety of procedures as part of their total evaluation activities, and the methods they use will vary depending on their assessment purposes.

In making decisions about second language instruction and learners, both qualitative and quantitative information are used. Qualitative information— that a student has an accent when she speaks English, for example — helps in planning special assignments: helping her improve her accent. Quantitative information — for example, that the average reading speed of first-year 10 students is 63 words per minute — may assist in selecting reading texts of appropriate average difficulty for second language learners.

There is not always a clear distinction between qualitative and quantitative information, but this is not necessarily a drawback. To say that a student's vocabulary in his second language is nativelike is unquestionably a qualitative statement. And it is clearly a quantitative assertion that he scored 604 on the TOEFL. But much evaluation information is expressed in terms that include both qualitative and quantitative aspects.

Second language evaluation involves the collection of both qualitative and quantitative information. In general, having a variety of types of information about teaching and learning can enhance the reliability of assessments and the validity of decision making.

All information, whether qualitative or quantitative, refers to characteristics of something: students or teachers, textbooks or videotapes, texts or realm, blackboards or ministries of education. For example, we can have information about a student's reading ability or a teacher's enthusiasm, about the instructional approach of a textbook or the topic of a videotape, about the complexity of a text or the familiarity of an object brought in-to the classroom, about the size of a blackboard or the political makeup of a department of education. It is necessary to be very clear about our information in order to avoid misunderstandings. Information about a student's knowledge of grammar is not the same thing as information about that student's ability to use the new language, even though teachers, students, and parents often interpret information about the first as if it were about the second.

Qualities of information

Regardless of what methods are used to collect information, one must always be concerned with the quality of the data used for evaluation. Two technical aspects of quality, i.e. reliability and validity, are to be discussed here. In the most general terms, reliability refers to consistency and stability, to freedom from nonsystematic fluctuation. Validity is the extent to which the information is relevant. Reliability and validity are both critical for judging the quality of qualitative information (for example, when observing student behavior in class or assessing their conversational skills) and quantitative information (for example, from tests or rating scales).

At the same time, there are additional matters to be considered when collecting information for evaluation. There is the practical side of gathering data.

Practicality

An obvious practical consideration when planning evaluation is cost. Some procedures, such as standardized tests, can be very expensive, and therefore their use is limited. Closely allied to financial cost is the administrative time required to collect information using certain procedures. Procedures that can be administered to large groups are obviously less time consuming than those that can be used only with individuals. Administration time can be especially important in schools with fixed class periods. Trying to schedule tests, questionnaires, rating scales, and so on, that require more time than a single class period can create major problems. Similar to administrative time is compilation time. All procedures for collecting information require time to transform results into a usable form.

Some procedures have demanding administrator qualifications, whereas others require no special talents or training. For most multiple-choice language tests, examiner qualifications pose no problems: Language teachers generally possess the qualities needed to administer such tests. A final practical attribute is acceptability. It can be extremely difficult to implement decisions based on information collected using procedures that students, their parents, or the community at large lack confidence in. Before using a particular procedure for making vital decisions, it is wise to determine whether it has community approval. If not, informing the community about the reasons for your choice may produce the desired approval. If it does not, you may wish to use a more acceptable procedure even though it may not be as desirable otherwise. Acceptability is sometimes called face validity. It is preferable to use the term acceptability, however, because this practical attribute does not share the technical characteristics associated with other types of validity. These practical aspects of information collection may be summarized as follows:

1. affordable cost of the method of information collection;
2. sufficient time in class to collect information using this method;
3. compilation time, i.e. enough time to score and interpret the information;
4. acceptability of the method to students.

A particular information-gathering procedure may be employed because of its practical attributes, but a procedure or an instrument shouldn’t be selected on the basis of practical qualities alone. The technical attributes of reliability and validity outweigh practicalities, and validity is the most crucial quality of all. Without validity, gathering information is, at best, a waste of time.

Reliability

Reliability is concerned with freedom from nonsystematic fluctuation. Information is subject to fluctuation, depending on who is observing. There are three general sources of unreliability: the first has to do with instability or nonsystematic fluctuation in the person or among the people collecting the information.

The second source of unreliability concerns the person about whom information is being collected. This was probably the source of unreliability with respect to a student’s writing problems. Depending upon when and what she is writing, different problems show up in quite unpredictable ways. This is called object-related or person-related reliability.

The third source of unreliability resides in the procedures used for collecting information. This is called instrument-related reliability.

None of the three sources of unreliability is intrinsically better or worse than the others. One always wants to get information that is as reliable as possible; fortunately, there are some very practical ways of doing so. Reliability related to the persons collecting information can be enhanced if they know exactly how to get the desired information and if they are well trained and experienced in the information collection procedures. Whenever possible, it is advisable to use more than a single observer, interviewer, or composition reader. Moreover, they should make their observations or do their interviewing or reading independently. If a number of independent people agree on their assessments, one can have much greater confidence in the reliability of that information.

Person- or object-related reliability can be enhanced by assessing on several occasions. This is especially advisable when human abilities or dualities are the object of assessment.. Using information about a student's performance or achievement collected on different occasions and using different procedures is highly recommended when making decisions about second language learners. It is best to avoid using students' performance on a single occasion (e.g., on a classroom test) as the sole basis for making decisions about them.

Instrument-related reliability can be improved by using a variety of methods of information collection. In this way, the bias or inaccuracy resulting from the use of one method will be offset by other methods; for example, second language learners with particular cultural backgrounds may find it difficult to demonstrate what they have learned if asked to do so in front of other students; they might be more comfortable doing so when alone with their teacher. Using only a method of evaluation that calls for performance in front of the entire class could lead to an unreliable estimate of their achievement. A more reliable procedure could be to ask students to keep a log of how much they study at home during one week.

A general strategy for thinking about how to enhance reliability is to begin by considering possible sources of unreliability. In other words, identify factors that, if not controlled for, are likely to result in inconsistent or variable estimates of performance. For example, unreliable estimates of performance may result from assessing student performance at times during the day or week when they are not at their best or are ill-prepared or when you are not at your best and cannot give your full and careful attention to assessing their performance. Unreliability can result from poor or inconsistent record keeping so that your assessment of a student's progress and your instructional planning for that student are based on inaccurately recalled information. Improving reliability of information for evaluation often involves reducing or averaging out nonsystematic fluctuations in assessors, objects, and instruments. In any given situation, not all sources of unreliability are equally threatening. One does the logical and prudent thing and puts major effort into reducing the greatest source of unreliability. For example, we need not be greatly concerned with rater-reliability when using multiple-choice tests, but we should be very concerned about this kind of reliability when assessing students' oral performance in an interview-type test. We must be concerned about person- and object-related reliability if we want to assess the communicative effectiveness of second language learners because their performance might fluctuate according to the time of day or their state of mind. In contrast, if we are assessing the linguistic complexity of a written text, object-related reliability is of little concern; that is, we need not be concerned that the text will change from moment to moment or from day to day.

Types of reliability and ways of enhancing reliability

Type of reliability	Ways of enhancing reliability
Rater reliability	Use experienced, trained raters Use more than one rater Raters should carry out their assessments independently
Person-related reliability	Assess on several occasions Assess when person is prepared and best able to perform well Ensure that person understands what is expected (that is, that instructions are clear)
Instrument-related reliability	Use different methods of assessment Use optimal assessment conditions, free from extraneous distractions Keep assessment conditions constant

In summary, the effects of unreliability on the quality of information you collect and the decisions you make can be serious, and, therefore, the reliability of your information should always be given careful attention.

Unfortunately, it is not actually possible to compute the true reliability of information or of the procedures used to collect data because we do not know the true state of affairs about the people or objects we are assessing, In practice, we can only estimate the reliability of our information or procedures. There are a number of different ways of doing this. Most of them yield indices that range from .00 for no reliability at all to 1.00 for perfect reliability. This is called a coefficient of reliability. For the moment, it is enough to know that:

- Reliability is a matter of degree and is usually expressed by indices ranging from .00 to 1.00
- Reliability can only be estimated and not truly calculated
- High reliability is desirable for any information used for evaluation purposes
- There are practical ways of enhancing reliability in classroom evaluation

Validity

Validity is the extent to which the information collected actually reflects the characteristic or attribute you want to know about.

There is an important relation between reliability and validity that you should be aware of— an assessment instrument or procedure (such as a test) can be only as valid as it is reliable. Worded differently, inconsistency in a measurement procedure reduces validity. An unreliable procedure or test is one that contains a lot of nonsystematic variation. In other words, the results of the procedure are influenced by sources of influence that are not those that the procedure or test is trying to assess. As a result, it produces inconsistent, erroneous, or unreliable results. Such non-systematic variation is like noise: it masks what you are really trying to measure. Since validity is the extent to which the information you get is the information you want, validity can be no greater than reliability. In fact, it will always be less. In short, a "noisy" instrument reduces validity.

Validity, like reliability, cannot be assessed directly. The reason why it cannot is the same. To assess the validity of information directly, you would have to be certain of the true state of affairs in order to compare it with the information you have collected. In the realm of human assessment, most of the qualities and attributes evaluators are interested in are not themselves subject to direct assessment. Thus, there is no direct way to know the true level of most human qualities or abilities that we are interested in. We have only indicators that allow us to make inferences about attributes of interest.

Since we cannot assess the validity of information about most human characteristics directly, we are forced to use indirect approaches to estimate the validity of our data and collection procedures. In the context of evaluation, in general assessment information is collected so that we can make sound educational decisions about students and instruction. If the information we have collected helps us to do so, we conclude that the information and the procedure for obtaining it have validity. Depending upon the kind of decisions we want to make, we employ different procedures to determine the validity of our assessment information and our methods of collecting it. There are three main procedures for doing this now.

Content Relevance

Content relevance is assessed logically by carefully and systematically examining whether the method and content of the assessment procedure are representative of the kinds of language skills you want to assess. Content relevance is important for classroom-based assessment because second language teachers often want to judge how students can perform in a range of situations or in certain types of situations when, in fact, it is not possible to assess student performance directly in the situations in question. Therefore, it is necessary to assess student performance in a restricted range of situations or in situations that are not exactly the ones we are interested in and then to generalize the results of this assessment to those situations we are most interested in. For example, an objective may be that students are able to converse with native speakers of the target language in situations outside of class that are typical of the age level of the students. Because the teacher cannot observe students conversing with native speakers of the same age in nonclass situations of the sort she is interested in, she might set up simulated conversations between her students and a native speaker that consist of prerecorded messages from a native speaker. The second language students' responses to these messages are tape-recorded and later assessed by the teacher. In effect, the teacher uses the information from these student conversations to infer how well they could actually converse with native speakers in authentic situations.

The teacher in this example could argue for the content relevance of her assessment procedure if she could demonstrate that the kinds of language skills called for in the simulated conversation are the same as those called for in authentic conversations involving native speakers. In other words, she should be able to show that the situations she observed her students in are representative of situations in which conversations could take place with native speakers.

Content relevance is important when devising classroom tests. It is also important when using standardized tests. In these cases, it is a question of whether the content of the test is representative of the kinds of language skills the teacher has taught and is interested in assessing. Let us take a placement example: students may be misplaced in specific classes because of a lack of content relevance in the placement test. If the content of the placement test does not accurately reflect the content of the classes that are offered, student performance on the test will not accurately predict their performance in those classes. Another way of saying this is that if there is little or no correspondence between the language skills required on the placement test and those needed to succeed in the available classes, we cannot accurately judge students' readiness for those classes.

The procedure for determining content relevance is not mathematical. Rather, it is logical and calls for good judgment. Content relevance can be characterized as high, moderate, or low, but it cannot be quantified.

Criterion-Relatedness

Criterion-relatedness is the extent to which information about some attribute or quality assessed by one method correlates with or is related to information about the same or a related quality assessed by a different method. For example, a teacher may want to use interesting texts in class because he is convinced that students learn more when the texts are intrinsically interesting. He collects information about students' preferences by showing them titles of texts and asking them to judge whether the passages that go with those titles are "very interesting," "somewhat interesting," "a bit dull," or "tedious." He records the interest categories selected by the students. Later he makes a record of the time that students spend, on average, working with each of the texts. Then he compares interest ratings with work times. If high interest ratings are associated with longer study times, the teacher may accept this as evidence for the validity of his procedure for assessing the interest value of texts for his students.

Another example of criterion-relatedness concerns testing. On the basis of their performance on a placement test, students are placed in beginning, intermediate, or advanced second language classes. Often, however, despite this procedure, it turns out that the students do not fit into the assigned class very well. Teachers find that they have a group of students with widely varying kinds and levels of second language ability and they do not do as well as expected. In these cases, the placement test clearly lacks criterion-relatedness: the ability to identify students with second language skills suitable for different levels of classroom instruction. The test was unable to predict the students' ability to handle instruction in specific classes.

Construct Validity

Construct validity is probably the most difficult to understand and the least useful for classroom-based evaluation although it can play an important role in judging the quality of standardized tests. A relatively uncomplicated example may help to explain what it is. Suppose a teacher of English wants to know how important each of her students considers the learning of English. She has them rate the importance to themselves on a scale ranging from "very important" to "no importance at all." Later, she compares the students' self-ratings with their achievement in the course. She finds, not surprisingly, that most of the students who indicated that knowing the language was important to them were among the best learners in the class. She concluded that the ratings were valid because students who are more motivated should in general be better learners. This illustrates the elements of construct validation:

1. You have information you want to validate.
2. You have a theory about how that information should relate to other
3. You can verify the predictions of your theory.

Let us now consider a more complex example. Suppose you want to assess the validity of a computer program that estimates the readability of written texts. You do not know how the program works, but you can get readability scores for a number of texts. You also have a theory about readability that states that readability is enhanced by explicit cohesion markers, lexical diversity, familiarity with the lexical items in the text, and familiarity with the content schemata. The theory also states that readability is unrelated to the number of clauses per sentence or the complexity of verb phrases in the text.

It would be relatively easy, although time consuming, to calculate indices of these linguistic variables for each of the texts. For example, you might construct an index of familiarity with schemata by letting students read the first half of the texts and then try to complete them. The amount of second-half content the students could guess would be an indication of familiarity with schemata. The last step would be to compare your information about the variables in the theory with the readability scores generated by the computer. The strongest evidence of construct validity would be afforded if the readability scores were positively related to indices of all the variables except density of clauses and complexity of verb phrases.

Construct validation is most useful when you do not know the exact content of the quality or attribute you want to assess, thereby ruling out the use of content validity. It is also useful when you have no well-defined or generally accepted criterion that could establish the criterion-relatedness of the assessment procedure.

Like reliability, the validity of assessment procedures can often be judged by identifying the possible factors that can invalidate them; in the case of test scores, for example, other factors besides second language ability might explain student performance, if there are a lot of other factors that could explain performance, especially pool performance, then the evaluation procedure probably has low validity as an indicator of second language proficiency. Improving validity is often a matter of eliminating, reducing, or otherwise taking into account these other factors. For example, poor performance may be due to a lack of understanding of what is expected, insufficient time to carry out the task, lack of interest in the activity or the possibility of performing the task in different ways that are equally valid but unforeseen by the evaluator. These possibilities can be reduced substantially if they are first seen as possible sources of contamination.

Thus three characteristics of information: practicality, reliability, and validity have been discussed. We pointed out that:

- These characteristics of information are vital for judging the quality of both quantitative and qualitative information.
- Reliable and valid procedures for collecting information are essential for sound educational decision making.
- Validity is the most important quality of information; furthermore, in classroom-based evaluation, content and criterion validity are desirable.
- Validity is related to (in fact, limited by) reliability.
- Validity and reliability are relative qualities, not absolute.
- Reliability and certain types of validity can be estimated statistically by an index that ranges from .00 to 1.0.
- Classroom teachers need to consider factors that can reduce the re-liability or validity of an evaluation procedure.
- There are practical ways of enhancing the reliability and validity of evaluation procedures that involve minimizing sources of unreliability or invalidity.

ГРАММАТИЧЕСКАЯ СЕМАНТИКА ЯЗЫКОВОЙ ЕДИНИЦЫ: ОРФОГРАФИЧЕСКИЙ АСПЕКТ (НА ПРИМЕРЕ СТРАДАТЕЛЬНОГО ПРИЧАСТИЯ И ИМЕНИ ПРИЛАГАТЕЛЬНОГО)

Колочкова О. В.
Казанский государственный университет

…а грамматику только тот знает, кто знает, как следовало по правилу сказать или написать то или другое слово, ту или другую фразу, которым живая область употребления (usus tyrannus) дала неправильную форму.

Н.В. Трунёв